Exploring Image Segmentation Methods to Segment Tumors by Training over a Dataset Marked by Skilled Professionals

BY

Varun Somani

SENIOR THESIS

Submitted in fulfillment of the requirements
for ECE 496/499:Senior Thesis
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Adviser:

Professor: D.A. Forsyth

Abstract

This research uses chest CT scan images of lung cancer patients to examine current methods in image segmentation in the context of tumor segmentation. Potential benefits of the research include faster processing and detection time for patients as well as allowing doctors to rapidly proceed with the requisite procedures.

We use both supervised and unsupervised methods to segment images. In terms of supervised methods, we use neural networks and SVMs with various “kernel tricks.”

In terms of unsupervised methods, we use K-means clustering and Otsu’s method.

Neural network gave the best result while other methods tended to have inferior performance. The results suggest that there is a possibility of further developing neural networks to conclusively solve the problem.

Key words: image segmentation, machine learning, lung CT, supervised learning, unsupervised learning

Contents

1. Introduction

The goal of the study is to examine existing segmentation methods and apply them in the context of segmenting tumors from medical images. Segmenting tumors accurately and manually is a time-consuming task for radiologists. It can take radiologists up to 10 minutes per patient, and there can be up to 20% variation between radiologists.[1] As such, extracting tumors automatically has a host of potential benefits such as improved diagnosis time and accuracy. In this paper, we survey methods to segment lung tumors in computed tomography (CT) data.

A CT scan is one of the most common, non-invasive procedures for the preliminary detection of for tumors. It is used to detect multiple types of cancers[2] such as lung, liver, kidney, etc. CT scans involve producing multiple cross sectional images of inside the body and combining them into a 3-D representation.

The data from a CT scan is published as a Digital Imaging and Communications in Medicine (DICOM) file. Each DICOM file is a stack of 2D images to form a 3-D representation of the inner body. The DICOM files are accompanied by an RTSTRUCT file that contains tumor delineations by experts that we use as ground truth when evaluating our segmented results.

We use multiple segmentation methods, both supervised and unsupervised. For supervised methods, we split our data into 80% training set and 20% test set. For unsupervised methods, we use only the test set to ensure comparability among the methods. We preprocessed the data before using any of the methods.

Challenges include an enormously unbalanced dataset, as most of the pixels corresponded to the non-tumor class. This is a problem because if a classifier predicts non-

tumor, it would be almost always correct. However, it would not be a very useful classifier, as it would never detect tumors.

2. Literature Review

As a precursor to this report, we refer primarily to the 2010 paper by Sharma et al.[3] The paper starts off by reporting on multiple means of obtaining medical data. It then continues to explain the various means of segmenting the obtained data, along with the advantages and disadvantages of each. The paper was primarily used as a reference for selection of techniques to test for this paper.

Next, the 2007 paper by Wirjadi[4] was used as a secondary reference. The paper describes various 3-D segmentation algorithms in detail. The author proposes various situations where some algorithms should be favored over others. As medical data is obtained as a 3D stack, this paper was used to further refine method selection from the previous paper.

Finally, we refer to the 2016 paper by Ray[1] and the 2013 paper by Gordillo et al.[5] to research previous results in the field. Ray achieved a dice score of 0.86 with his results; however, his methods relied on drawing a bounding box specifically around the tumor instead of the entire lung. Gordillo et al. explored segmentation methods on a brain tumor dataset and found that thresholding methods tend to work under limited conditions. They found that region-based or model based approaches work under semi-automated conditions, such as when radiologists could mark a seed point. Neural networks yield results similar to Ray's paper.

3. Method

3.1 Preprocessing

The dataset was obtained from The Cancer Imaging Archive.[6] The dataset consisted of both the scans and the ground truth as delineated tumors. It consisted of 100 patients. There were 134 images per patient, each 512x512 pixels. Each patient's file consisted of a full body CT scan. As the first step, we extracted all the slices from the 25th to the 100th for each patient, as these were the most representative of the lung location

Then, for each image in the extracted stack, we crop the middle 256x256 pixels as these were the most representative of the location of the lung and the tumor.

Finally, for each patient, we extract three contiguous images in their respective 3D stack that contained the most tumor pixels. We do this by examining the corresponding ground truth image. This step of pre-processing simulates the radiologist selecting the most relevant image in a CT scan.

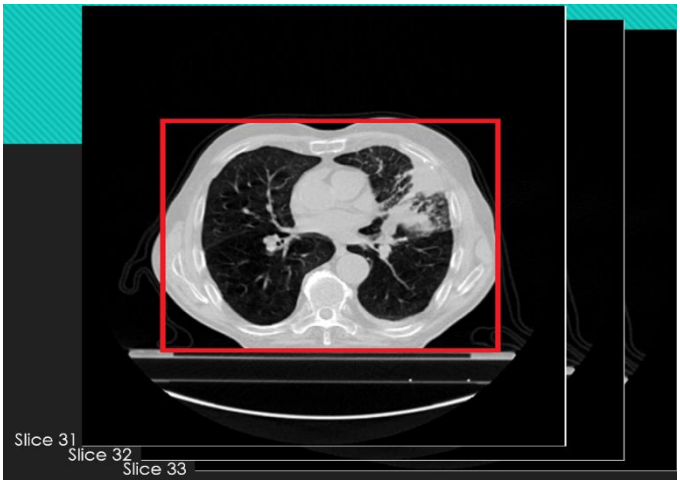

Figure 1 Stacked 2D slices to form 3D image

We then combine those three contiguous images in a single stack to form a new 3D image that forms the basis of our training and test data. Figure 1 shows one of the resulting stacks. The red region is the selected region in each slice and the rest of the image is discarded.

3.2 Unsupervised Methods

Unsupervised learning methods used in this report are global thresholding with Otsu's method and clustering with K-means. We use both clustering and thresholding to explore multiple approaches to the problem.

3.2.1 Otsu's Method

Otsu's method[7] is a simple method which involves computing a threshold that minimizes the intra-class variance. The resulting threshold is then used to segment the image. Equation 1 represents the weight sum of variances of the two classes. The terms, w_0 and w_1 are probabilities of the two classes separated by threshold t and σ_n^2 represents the variance of the classes. We iteratively search for the threshold that minimizes the weighted intra-class variance.

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (1)$$

3.2.2 K-Means Clustering

K-means clustering uses an iterative refinement that continues until convergence or the max iteration step. The goal is to compute cluster centers that best minimize inter-cluster squared distance. We use K-means to divide the 3-D representation into three clusters. The cluster that maximally tests against the ground truth is the cluster we consider to represent the

tumor. We assume only one of the clusters corresponds to the tumor, as we assume the other clusters correspond to either the lung or the background.

We use K-means on each 3-D stack in the test set. Afterwards, we generate sixteen 16x16x3 grids per each 3D stack to generate multiple local centers instead of three global centers. This was to test local centers and global centers to see if there was a difference in the results.

3.3 Supervised Methods

The protocol for supervised methods was 70% training, 20% validation and 10% test. We used dithering for neural networks by random rotation and scaling to generate more data. Rotation was between +5 and -5 degree arc, and scaling was 95% to 105%. All resulting images had dimensions 256x256 pixels.

3.3.1 SVM

Support vector machines are classifiers that attempt to learn a hyperplane to discriminate between the classes. In this problem, we are using the SVM as a pixel-level classifier. To train the SVM, we create a 1x27 vector of the pixel and its neighbors in 3D. We then augment the vector with the absolute x, y, z location of the pixel. We augment the pixel because tumors are more common in some locations than in others. Each such vector corresponds to a training sample. We under-sample the non-tumor class to combat the imbalance in the dataset, so we have 50% tumor and 50% non-tumor samples.

3.3.1.1 Linear Kernel

A linear kernel performs no dimensionality change on the input data and attempts to compute the hyperplane in the same dimension as the input data.

3.3.1.2 Radial Basis Function Kernel

A radial basis function (RBF) or Gaussian kernel maps the input as shown in equation 2. This is done under the assumption that the data may be linearly separable in a higher dimension.

$$K(x, x') = \exp(-\frac{||x-x'||^2}{2\sigma^2}) \quad (2)$$

3.3.2 Neural Network

A neural network is a computational system composed of multiple layers consisting of weighted interconnections. We use the SegNet-Basic architecture,[8] a network originally designed to segment scenery into 11 classes. We use the implementation provided by Sungjoon Choi[9] for Tensorflow to train on the dataset. We modify the implementation to work on our dataset and add a higher cost to predicting the non-tumor class. A higher cost counters the data imbalance by forcing the classifier to recognize the smaller tumor pixel class. It was trained for 16 hours on an Nvidia GTX 1080.

4. Description of Research Results

The results are summarized in table 1. Raw accuracy is not a representative measure of the results, as there is heavy imbalance in the dataset. The more relevant score is the recall, as identifying all the tumor pixels is more important than identifying a non-tumor pixel correctly.

Table 1: Results

Classifier	Accuracy	Precision	Recall	Area Under Curve
Otsu's Method	92.2%	2.4%	11.4%	0.56
Gridded K-Means	71.9%	2.4%	10.2%	0.48
K-Means Full	75.6%	0.28%	5.6%	0.51
SVM-Linear	64.4%	53.6%	3.3%	0.57
SVM-RBF	98.6%	0.03%	0.2%	0.51
Neural Network Original	97.2%	16.3%	36.7%	0.64
Neural Network 2x Dither	97.5%	16.8%	38.7%	0.66
Neural Network 4x Dither	97.7%	17.4%	43.4%	0.68

4.1 Receiver Operating Curves

Receiver operating curves of the methods above are shown in figures 2 through 7. All the techniques have an AUC of around 0.5, except for the neural networks, which were significantly higher. Dithering improved results, as can be seen from the dithered ROC curve.

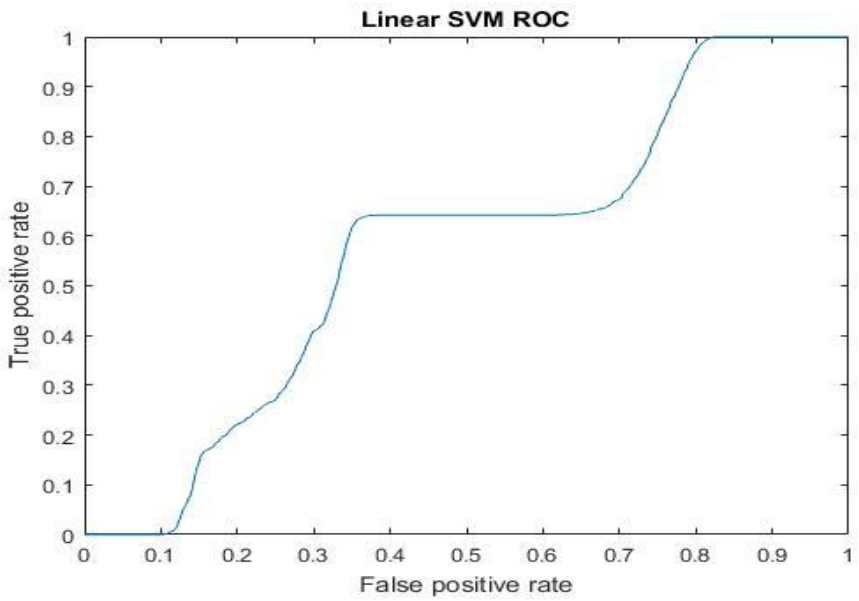

Figure 2 Receiver Operating Curve for SVM with linear kernel

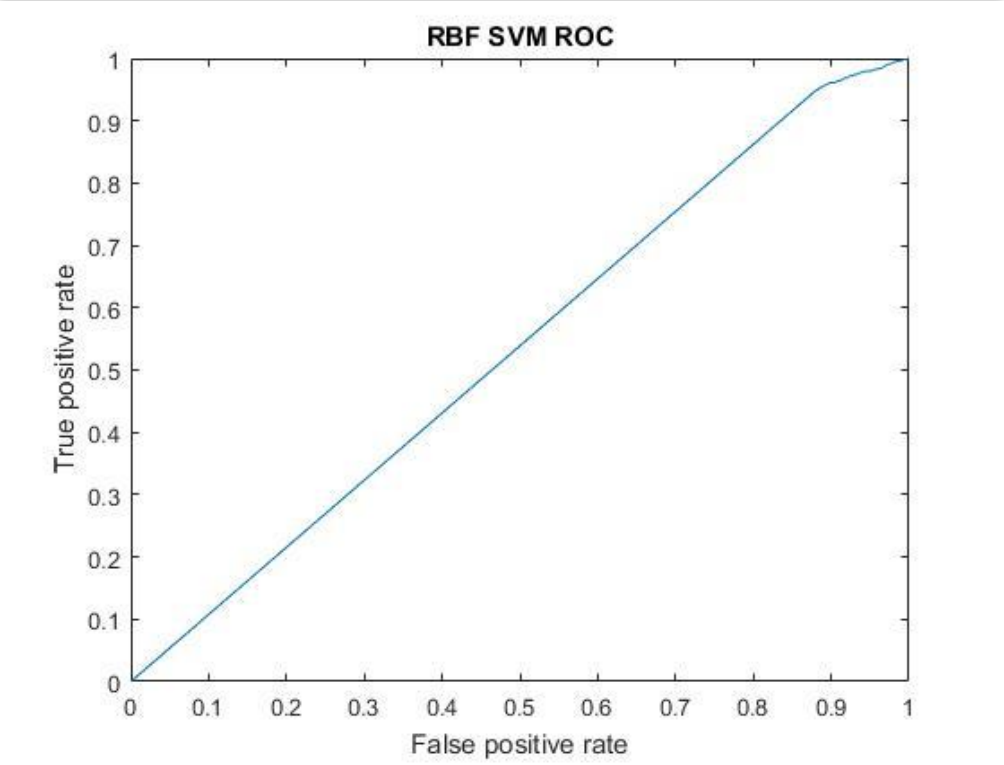

Figure 3 Receiver Operating Curve for SVM with RBF kernel

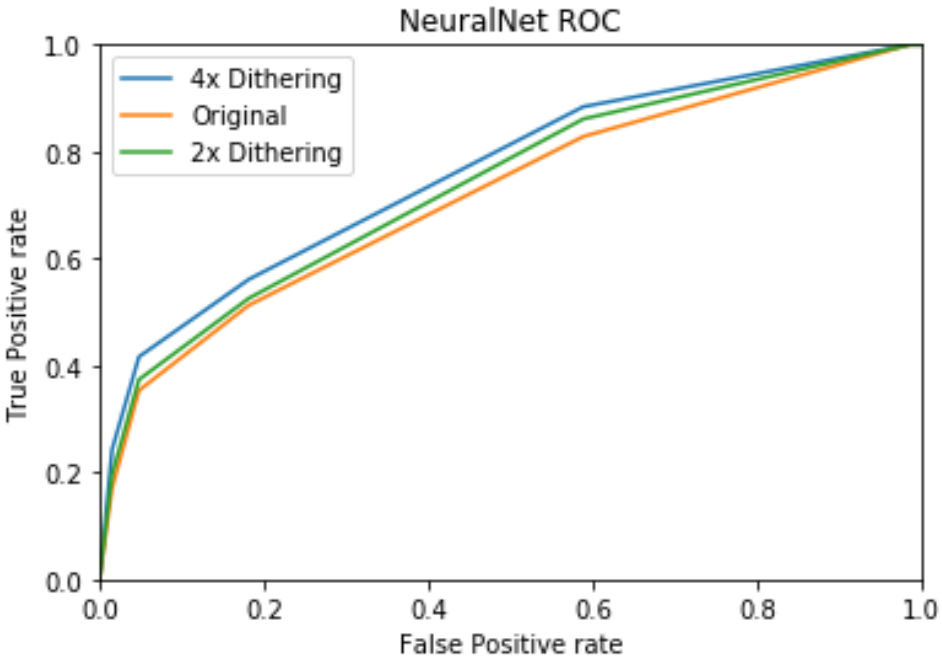

Figure 4 Receiver Operating Curve for SegNet-Basic

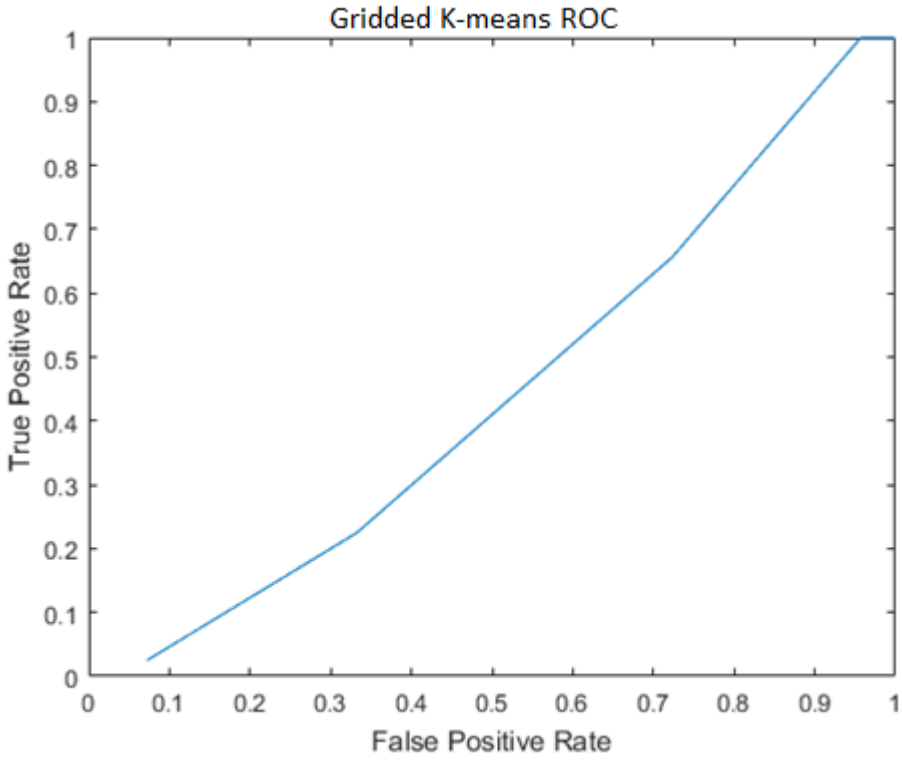

Figure 5 Receiver Operating Curve for K-means Gridded

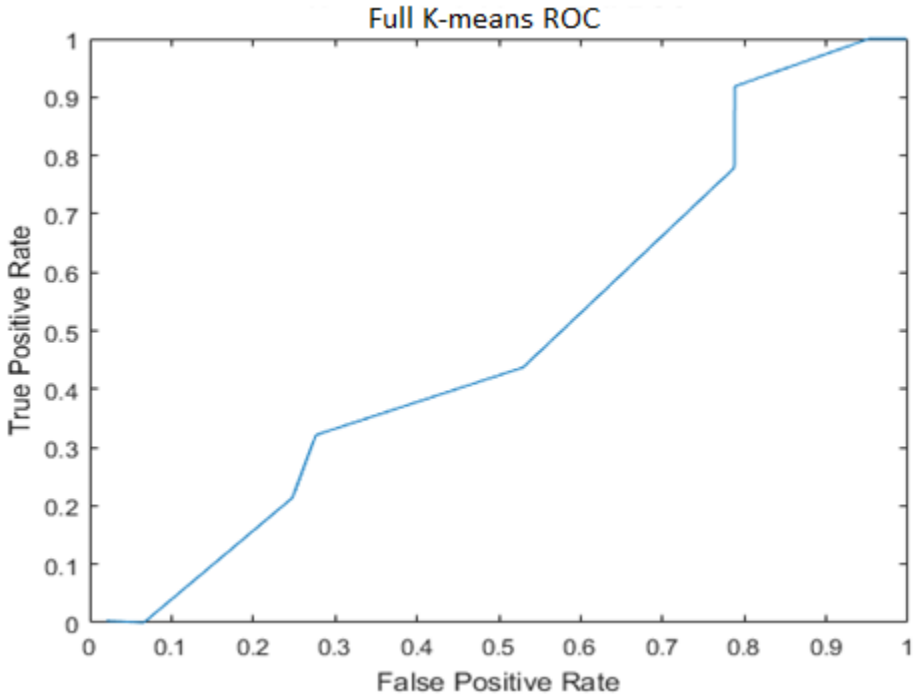

Figure 6 Receiver Operating Curve for K-means Full

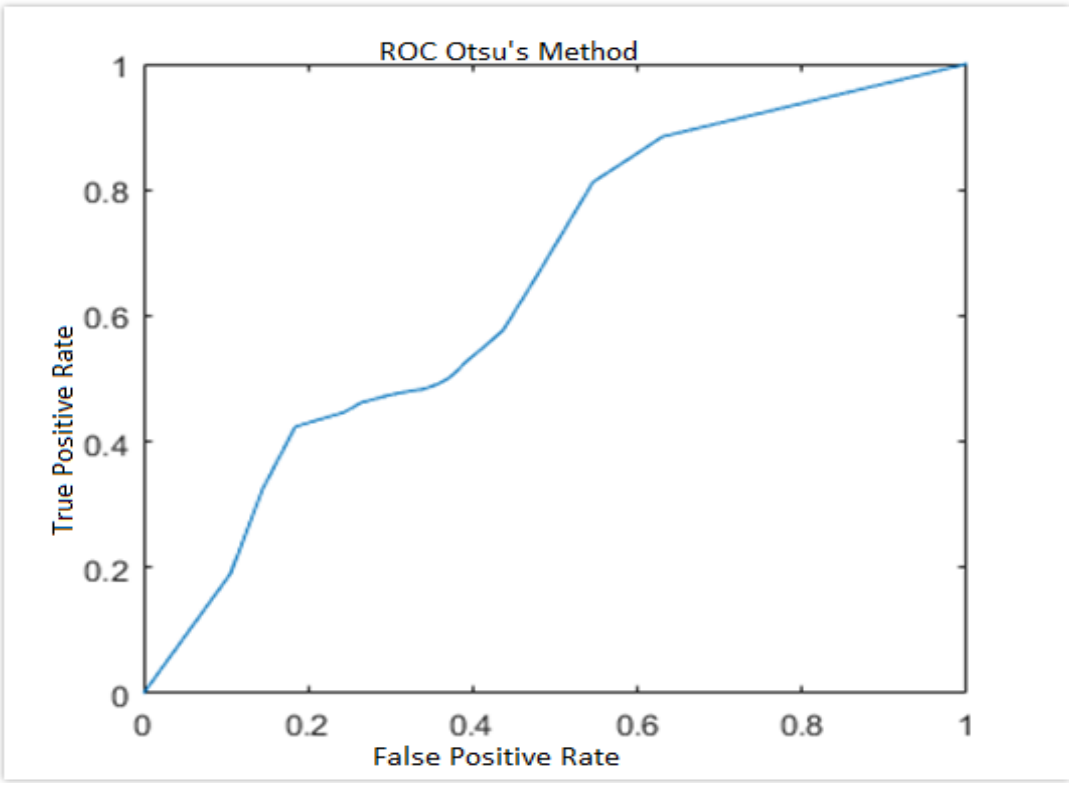

Figure 7 Receiver Operating Curve for Otsu's method

5. Conclusion

Based on the numbers obtained in Table 1 and the resulting ROC graphs, it can be fairly concluded that neural networks provide the best results for such a task. The neural network had the best resulting receiver operating curve and the best recall and accuracy rates.

The neural net architecture was selected because of the relatively lower computational requirements. However, there are other architectures that may produce better results, such as the Fully Convolutional Network (FCN) U-Net[10] which works with volumetric data. Testing these networks was beyond the scope of a senior thesis, but may provide avenues for further research. Further, increased data diversity and volume will also improve the results.

References

1. A. Ray, "Lung Tumor Segmentation via Fully Convolutional Neural Networks", Available: http://cs231n.stanford.edu/reports/2016/pdfs/302_Report.pdf , 2016.
2. "Body CT (CAT Scan)", Radiologyinfo.org, 2017. [Online]. Available: https://www.radiologyinfo.org/en/info.cfm?pg=bodyct. [Accessed: 07- Feb- 2017].
3. N. Sharma, A. Ray, K. Shukla, S. Sharma, S. Pradhan, A. Srivastva and L. Aggarwal, "Automated medical image segmentation techniques", Journal of Medical Physics, vol. 35, no. 1, p. 3, 2010.
4. O. Wirjadi, “Survey of 3D image segmentation methods”, *Tech. Rep. 123*, Fraunhofer ITWM, Kaiserslautern, Germany, 2007.
5. N. Gordillo, E. Montseny and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation", Magnetic Resonance Imaging, vol. 31, no. 8, pp. 1426-1438, 2013.
6. Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Cavalho, S., … Lambin, P. (2014, “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature Communications. Nature Publishing Group”. http://doi.org/10.1038/ncomms5006 2015.
7. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.
8. V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1-1, 2017.
9. S. Choi, “ Tensorflow-101 ”, 2016 , [Online]. Available https://github.com/sjchoi86/Tensorflow101/ .
10. O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. LNCS, vol. 9351, pp. 234–241. Springer 2015.